《心理学报》论文自检报告

请作者填写以下内容，粘贴在稿件的首页。

1. 请以“研究亮点”的形式列出最多三条本研究的创新性贡献，总共不超过 200 字。
《心理学报》的目标是发表“既科学优秀，又具有广泛兴趣和意义” (be both scientifically excellent and of particularly broad interest and significance)的心理学前沿研究。如果您的研究只有小修小补的贡献，没有尝试开创新的研究领域(new areas of inquiry)或提出独到见解和创新视角(unique and innovative perspectives)，特别纯粹只是研究没有明确心理学问题的算法或技术的工作，这类研究被本刊接受的机会小，建议另投他刊。
答：证明了视觉加工中的结构分析系统的对应 ERP 成分

2. 作者已经投稿或发表的文章中是否采用了与本研究相同的数据？如果是，请把文章附上审查。(我们不赞成作者用同一数据发表多篇变量相同的文章，也不赞成将一系列的相关研究拆成多个研究来发表的做法。)
答：否

3. 管理、临床、人格和社会等领域仅有自我报告(问卷法)的非实验非干预研究，需要检查数据是否存在共同方法偏差(common method bias)。为控制或证明这种偏差不会影响研究结论的效度，你使用了什么方法？采取了哪些措施？(共同方法偏差的有关文献可参见：http://journal.psych.ac.cn/xlkxjz/CN/abstract/abstract894.shtml)基于横断数据，仅有自我报告，仅仅在方便样本中施测，这样的研究数据易取得，但通常创新性价值不大，被本刊接受的机会小。
答：不是问卷

4. 是否报告并分析了效果量(effect sizes；如：*t* 检验：Cohen’s *d*；方差分析：η^2 或 η_p^2；标准化回归系数)？(很多研究只是机械地报告了效果量，但没有做必要的分析或说明，如效果量是大中小？有什么理论意义或应用意义？)。(在 google 中搜索“effect size calculator”，可搜到许多计算方便的 APP。效应量的有关解释，中文可参考：http://journal.psych.ac.cn/xlkxjz/CN/abstract/abstract1150.shtml；英文可参看：http://www.uccs.edu/lbecker/effect-size.html
是否报告统计分析的 95% CI？(如，差异的 95% CI；相关/回归系数的 95% CI)置信区间的有关计算和绘图可参考 https://thenewstatistics.com/itns/esci/)
答：是 是

5. 请写出计划的样本量，实际的样本量。如果二者有差别，请写出理由。以往心理学研究中普遍存在样本量不足导致的低统计功效(power)问题，我们建议在论文的方法部分解释您计算及认定样本量的依据。应该以有一定依据的效果量(effect size)、期望的功效来确定样本量，并报告计算用软件或程序。样本量计划的理由和做法可参考 https://osf.io/5awp4/
答：计划样本量 20，实际样本量 11（1 名被试在晚上完成脑电实验，3 名被试未理解实验要求，5 名被试无效试次多于 25%）

6. 假设检验中，如果是零假设显著性检验(NHST)，需报告精确 *p* 值而不是 *p* 的区间(小于 0.001 的报告区间，其他报告精确 *p* 值)。你的论文是否符合该项要求？如果是贝叶斯因素，是否已报告其对先验分布假定的敏感性？
答：符合

7. 为保证论文中数据报告的完备性，统计分析中如果剔除了部分数据，是否在文中报告？原因是什么？包含这部分数据的统计结果如何变化？统计分析中是如何处理缺失数据的？使用量表时是否删除了其中的个别题目？原因是什么？如果包含这部分题目，统计结果会如何变化？是否有测量的项目或者变量没有报告？原因是什么？请写出在论文中的位置。
答：非问卷类研究，不涉及上述问题

8. 研究用到的未经过同行评议和审查的实验材料、量表或问卷，是否附在文件的末尾以供审查？如果没有，请写出理由。如果该文发表，您是否愿意公开这些材料与其他研究者共享？
答：文中提供了部分实验材料，如果发表愿意公开材料

9. 本刊要求作者提供原始数据，请在以下 3 种里选择一种打√：
a) 投稿后将数据发至编辑部邮箱 (√)
b) 数据可以从如下链接中获得

( )

c) 原 始 数 据 和 程 序 已 在 心 理 科 学 数 据 银 行 (https://psych.scidb.cn/) 上 分 享 ( )

d)如不能提供，请说明理由或提供有关证明。

10. 您的研究是否是临床干预或实验室实验？ 是(√) 否( )

如果是，请提供预注册登记号 ______________________。

如果没有，请说明原因____________脑电实验______________________________。

注：临床干预或实验室实验，建议在收集数据前预注册(pre-register)。也鼓励其他实验研究预注册。预注册要求写出所有的研究假设及其支持，以及实验/干预的详细过程和步骤。本期刊的预注册网站是 https://os.psych.ac.cn/preregister (使用说明书见本刊网站"下载中心")或 https://osf.io/ 或 https://aspredicted.org/。如果您的研究有预注册，会显著增加被录用的机会。预注册的重要性可参考 https://osf.io/5awp4/

11. 您的研究如果用到了人类或动物被试，是否得到所在单位伦理委员会的批准？如果是，请把扫描版发至编辑部邮箱。如果否，请说明理由。

答：否，工作单位没有成立伦理委员会

12. 是否依据编辑部网站发布的"英文摘要写作注意事项"撰写 400~500 个单词的英文大摘要？英文题目和摘要是否已请英语好的专业人士把关或者已送专业 SCI/SSCI 论文编辑公司修改润色？

答：是全英文稿件

13. 如果第一作者是学生，请导师单独给编辑部(xuebao@psych.ac.cn)发邮件，说明已阅读本文并认真把关。是否已提醒导师给编辑部发邮件？(编辑部收到导师邮件后才会考虑进入稿件处理流程)

答：非学生

14. 请到编辑部网站首页右侧"下载中心"下载并填写"稿件不涉密证明"，加盖通讯作者单位的保密办公章，把扫描件发至编辑部邮箱(xuebao@psych.ac.cn)。如没有保密办公章，请加盖通讯作者的单位公章。是否已发邮件？

答：是

Form analysis system: An EEG study of object, word, and Greeble recognition

Abstract:

Objectives: The form analysis system efficiently conceptualizes how object recognition is encoded in a frame-and-fill model. However, little is known about the neural basis of the form system. The present study aimed to narrow this gap using EEG. **Methods:** Participants were instructed to passively view six types of images: geometric shapes, animal headless bodies, plants, Chinese words, English words, and Greebles. **Result:** Shared negativity waves in the occipital lobe from 100 ms to 200 ms were observed across the three object domains, including geometric figures, animal bodies, and plants, but not observed in Chinese characters, English words, or Greebles.

Conclusion: The form analysis system was engaged with geometries, bodies, and plants, but not with words or faces. These results suggest that stimuli holding the medial axis structure can induce similar negativity waves in the human brain.

Our study sheds new light into the human visual system, revealing a form analysis system existed. Understanding the neural patterns of the form analysis system enhances our comprehension of visual object recognition. It could inform advancements not only in human visual cognition research but also in machine visual fields.

Keywords: Form analysis system; Objects; Words; Faces; EEG

“Seeing is a constructive process in which the brain responds in parallel to many different features of the visual scene and attempts to combine them into meaningful

wholes, using its past experience as a guide."

--Francis Crick

There is clearly a continuum between visual and semantic processing: dissimilar visual shapes may be identified as the same category, while different linguistic labels may represent similar visual appearances. The inferior temporal (IT) cortex, traditionally associated with visual object recognition, is also known to encode semantic dimensions (Khaligh-Razavi, Kriegeskorte., 2014). This highlights the challenge of distinguishing between visual and semantic categorical representations. However, converging evidence shows that visual encoding begins in early visual cortices (e.g. V1, V2, V3), while semantical information is processed in higher-level visual areas (e.g. IT). This suggests that visual encoding is a more fundamental process preceding the semantic process. Numerous theories have aimed to identify the fundamental visual elements of object recognition, generally falling into three main classes: global Gestalt perception (Rennig et al., 2015), shared converging features (Coutanche & Thompson-Schill., 2014), and statistical Bayesian inference (Erdogan & Jacobs., 2017). Recently, a new account has been proposed: the form analysis system (Spelke, 2022). Although many studies have demonstrated the importance of forms in object categorization, few hold the belief that the form analysis system is innate until a proposal suggested infants and adults are predisposed to recognize and categorize objects based on shape structure delineations that capture the form characteristic of plants and animals (Spelke., 2022).

Ever since Blum introduced the medial axis transform (MAT) in 1973, form

skeletal representation of visual shape has played a prominent role in theories of visual shape. Recently, a skeleton-based sign language recognition and generation framework has been proposed to support bidirectional communication between deaf and hearing people (Xiao, Qin, & Yin., 2020). The communicative power of skeleton information is such that these skeleton-based signs can be effortlessly identified without any language. Beyond efficient processing, skeleton-varied objects can also be encoded immediately. Using high-density electroencephalography (EEG), researchers found that the classification of animate and inanimate categories occurred around 60 milliseconds after stimuli onset, and some individual exemplar comparisons could be identified even earlier, around 40 milliseconds (Gurariy, Mruczek, Snow, & Caplovitz., 2022). Why are we so sensitive to skeleton-based information? Is there a special system analyzing these skeleton properties across various categories? By encapsulating evidence from different domains, including objects, diverse writing systems, and social face stimuli, the present study, using EEG, aimed to investigate whether or not the form analysis system exists.

Form Analysis in objects

Objects in the physical world can be abstractly represented as geometry. The human mind is equipped with a form analysis system to analyze surrounding visual objects in terms of geometric shapes (Wilder et al., 2019). When adults were instructed to tap on a figure anywhere they wanted, they tended to tap on the medial axis, the form of these geometries (Psotka,1978; Firestone, Scholl, 2014; Ayzenberg et al., 2019). This effect persists even when small perturbations in the shape contours

(Feldman, Singh, 2006; Ayzenberg et al., 2019). However, when asked to predict the tapping results, few adults developed an awareness of the geometric forms: one-third chose the chance model, while only 3% correctly identified the medial axis (Firestone, Scholl, 2014).

Though the form analysis system may not be consciously recognized, it is a basic element shared across all ages and cultures. Researchers compared geometric knowledge among adults and children (aged 6-10 years) from urban communities (in the United States or France) and an isolated community in the Brazilian Amazon: the Mundurucu (Dehaene, Izard, Pica, & Spelke., 2006; Izard, Pica, & Spelke, 2021). When asked to locate geometric deviants in panels of six forms with variable orientations, adults and children from both urban and rural communities showed strong similarities in their geometric intuitions: difficulty problems shared among all ages and in both cultures, suggesting that form analysis is a universal aspect of the human mind.

The form analysis system not only exists in literacy and illiteracy populations but is also shared among newborn humans and non-human animals. Newborns' preferential looking between pairs of stimuli varying in real size and viewing distance was only determined by retinal size, suggesting that an invariant form representation can be abstracted from the size-changing visual appearance on the retina (Slater, mattock, & Brown., 1990). In a follow-up experiment, newborns were desensitized to changes in distance and retinal size during familiarization trials. Subsequently, they strongly preferred a different-sized object over the familiar one, indicating that the

form-invariant representation not only abstracted size constancy on the retina but also captured the real objects' appearance across varied retinal sizes. These results suggest that newborn visual systems can begin building form-invariant object representations from the onset of visual object experience. Evidence from newborn animals revealed similar results. Using the imprinting response, Wood (2013) tested newborn chickens' object recognition abilities without training. During the input phase, the imprinted object was displayed from a single 60° viewpoint range, balanced on the left and right display walls. The test phase examined whether newborn chickens could recognize virtual objects across changes in viewpoint. Distinguishing between these objects from novel viewpoints requires an invariant form representation that can generalize across large, novel, and complex changes in the object’s appearance. The results showed that, from the onset of visual experience, newborn chickens generate form-invariant object representations from changing viewpoints.

Furthermore, the form analysis system is not limited to the geometries domain but is also presented in the bodies and plants realms. Gunnar Johansson (1950) combined light bulbs on an experimenter's joints and presented adult participants with these biological moving light-dots displayed in the dark. When the experimenter was walking, participants immediately perceived these spatial separate point-lights as a human in motion. This finding revealed that participants processed these insolated light dots as a solid body consisting of limbs and joints, consistent with the form skeleton analysis system. Like bodies, humans are also sensitive to the biological motion of plants in point-light displays (Cutting, 1982). This sensitivity appears to be

well-prepared since infancy. Researchers designed experiments with plants that had either a natural or unnatural growing structure, with or without leaves (Sarmiento & Spelke, *in process*). They found that children took longer to touch the natural growing structures and those with leaves, indicating an innate sensitivity to biologically relevant forms.

Form Analysis in different writing systems

Beyond the ancient objects shared with animals, humans have developed special symbols: words. Reading is one of the most well-practiced abilities for people in modern societies. With extensive practice, the huamn mnid raed wrods as a wlohe (Grainger & Whitney., 2004). Despite the letter position in the last sentence being transposed or jumbled, it is still easy to understand the meaning. This phenomenon is known as the "transposed-letter effect".

Converging evidence suggests that Indo-European languages are tolerant of imprecise letter positions in word identification. Researchers compared the electrophysiological response to four types of conditions: transposed-letter nonword-word pairs (e.g. “wlohe-whole”), transposed-letter word-word pairs (e.g. “calm-clam”), substituted-letter nonword-word pairs (e.g. “sitinar-similar”), substituted-letter word-word pairs (e.g. “soft-salt”). They found that only the substitution-letter nonword-word pairs elicited a more negative waveform between 150 and 250 ms, characterized as an N250, which is sensitive to form-level processing (Dun˜abeitia et al., 2009). These results suggest that transposed-letter nonwords more easily failed to be recognized as nonwords, as demonstrated in a behavioral visual

lexical decision task (Andrews, 1996; Chambers, 1979). Similarly, priming studies show that relative to a substitution prime (e.g. sedlice-SERVICE), a transposed-letter prime (e.g. sevrice-SERVICE) speeds up the recognition rate of a target word (Schoonbaert & Grainger, 2004). This transposed-letter effect extends to cases in which the transposition crosses a syllable boundary (e.g. snawdcih-sandwich; Guerrera & Forster, 2008) and to more extreme modifications (e.g. R34D1NG W0RD5 W1TH NUMB3R5; Perea, Dun˜abeitia, & Carreiras., 2008). These findings suggest a high degree of perceptual similarity between word and nonword stimuli that comprise the same letters in different positions. Since each letter has its unique form, this perceptual similarity can be explained by the form analysis system.

In most writing systems, words are written in a specific direction and orientation. Besides letter position, orientation is also crucial for word perception. In Chinese writing systems, characters are composed of the same stroke pattern but in different orientations can represent different meanings (e.g., “甲” means excellent, “由” means by, “陪" means accompanying, and “部” means ministry; Zhang, Ni, & Li., 2020). For young Chinese readers, these characters can be visually confusing. In the English writing systems, some letters share geometric forms that are mirror images of each other, such as “b” and “d”, “p” and “q” (Freud, Behrmann, & Snow., 2020). Furthermore, the inversion effect is another aspect of orientation (Diamond & Carey, 1986; Gauthier, Williams, Tarr, & Tanaka, 1998; Valentine, 1988; Yin, 1969). Researchers investigated the inversion effect in Chinese character processing (Kao, Chen, & Chen., 2010). They found that the proportion of correct responses for

matching real characters significantly reduced when the characters were turned upside down (96% vs 93%). However, no significant difference was observed in performance for upright and inverted non-characters (95% vs 94%). Using a novel visual perspective-taking task, Surtees and collaborators (2012) instructed child and adult participants to make judgments about the appearance of numerals, such as 6 and 9, which appear different when inverted. When a numeral is presented on the wall, it appears the same to both the self and avatar; while presented on the table, the stimulus is viewed inverted. They found that all participants had more difficulty recognizing a character when the stimuli were inverted than when presented upright.

By acquiring sensitivity to characters' orientations, the form analysis system can help children recognize mirrored and inverted characters more effectively.

Form Analysis in face

Face recognition and visual word recognition are both examples of expert perceptual skills acquired through years of practice. Holistic processing, in which individuals have difficulty ignoring irrelevant face information but focus more on the selected part of a face, is a hallmark of form analysis in face recognition (Tanaka and Farah, 1993; Young, Hellawell, & Hay, 1987). Both the face inversion effect and the part-whole effect underlies the holistic mechanism. The inversion effect (Thompson, 1980) illustrates that when a face's eyes and mouth are inverted while the rest of the face remains upright, the face appears distorted only when viewed upright but not when inverted. This vulnerability to orientation suggests that these facial features constitute the core form structure of facial identification. On the other hand, the

part-whole effect (Tanaka & Simonyi, 2016) demonstrates that individuals are better at recognizing isolated facial features, such as the eyes, mouth, or nose, when presented within the original face rather than in isolation or within a different face. The whole advantage suggests that face parts are not processed as standalone features but rather as integral forms of the whole face.

Tarr and his colleagues designed a new type of artificial object called “Greeble”, which shares a facial configuration with a geometric body (Gauthier & Tarr., 1997). This was intended as ideal control stimuli for face studies. The logic behind studies using Greebles is to investigate whether non-face objects can produce similar behavioral or neural effects as face stimuli. Researchers found that Greebles were processed more holistically by experts than by novices (Gauthier & Tarr, 1997, 2002), suggesting that Greebles could be perceived as face-like stimuli at a behavioral level. Greeble processing also shared a neural basis with face stimuli. The Fusiform Face Area (FFA), a brain region specialized for faces, was activated when subjects passively viewed Greebles (Gauthier & Tarr, 2002; Gauthier et al., 1999). Additionally, the face-specified ERP component N170 was observed when participants looked at the Greebles stimuli (Rossion et al., 2000). Furthermore, pathological evidence suggested that Greebles are processed similarly to faces and share similar mechanisms (Gauthier et al. 1999; but see Gauthier, Behrmann, & Tarr., 2004).

We used Greebles as face-like stimuli in this study for two reasons. First, studies involving Greeble provide converging evidence that face-specific effects can be

obtained with visually similar non-face objects. Second, the facial features of Greebles can be removed, allowing us to compare social forms with geometric features more effectively.

Form Analysis in Brain

The human ventral temporal cortex (VTC) is a key structure in high-level visual processing, including object recognition (Grill-Spector, Kourtzi, & Kanwisher, 2001), reading (Cohen, et al., 2000; Wandell, Rauschecker, & Yeatman, 2012) and face perception (Kanwisher, McDermott, & Chun, 1997). Some research suggests that the form analysis system exists in the infratemporal cortex. Bao and his colleagues (2020) divided the inferotemporal (IT) cortex, the core brain area responsible for object recognition, into four networks. These not only included the established face and body systems but also introduced two new networks: the NML network and the stubby network. However, these four networks cover only about 53% of the IT cortex, leaving many areas unexplored. Another crucial area for object recognition is the lateral occipital cortex (LO). Ayzenberg and his colleagues (2022) used fMRI revealing that the skeletal model explained significant unique variance in the response of the LO. However, this research used man-made artificial stimuli. To our knowledge, no study has explored the neural basis of the form analysis system using natural stimuli, such as animal skeletons and bare trees, or in other domains where the form analysis system exists.

To sum up, this study aimed to identify the brain areas activated when participants passively viewed various stimuli, including objects (such as geometric

figures, animal bodies, and plants), Chinese words, English words, Greebles, and the form counterparts of all these stimuli as the experimental group.

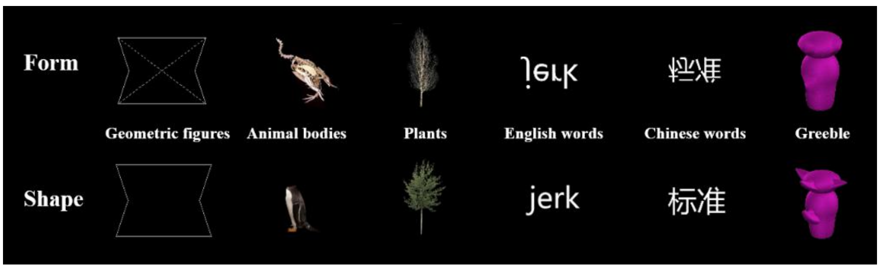
Materials and Methods

Participants

Twenty health subjects were recruited from the University Community. Nine participants were excluded during data analysis: one participated in the evening, three misunderstood the task, and five had more than 25% invalid trials. The remaining eleven undergraduate and graduate students (7 female, mean age = 22.2±2.61 years, range = 18.7-25.7) were included in the data analysis. All participants were right-handed, had normal or corrected-to-normal vision, and had no history of neurological/psychiatric disorders, or reading/learning difficulties. This study was approved by the Ethics Committee of our University.

Materials and Procedure

The stimuli (samples shown in **Figure 1**) consisted of black-scale images from six different object categories (12 exemplars per category): geometric figures, animal bodies without heads (in a neutral, standing position), plants, Chinese words (from the List of core vocabulary in Mandarin Chinese, published by Cambridge Assessment International Education, 2023), English words (from the British Lexical Project database; Keuleers, Lacey, Rastle, & Brysbaert., 2011), and Greebles (materials reproduced from Professor Tarr's lab; Bukach et al., 2012). Each of the six object categories was divided into two levels: shape and form. We were particularly



interested in the differences between these levels. To clarify this distinction, we defined the more familiar stimuli with outline contours as the shape level (see the bottom row in Fig. 1), and the designed stimuli with inner skeletons as the form level (see the top row in Fig. 1).

Figure 1. Samples of stimuli.

For the first three categories—geometric figures, animal bodies without heads, and plants—the concept of the form level was clear. Geometric figures contained the medial skeletal axis, animal bodies included their skeletons, and plants were represented as skeletal bare trees without leaves. All stimuli in these three categories at the form level incorporated skeleton information, making the difference between shape and form levels meaningful. In contrast, the last three categories (Chinese words, English words, and Greebles) were man-made and were interpreted based on their typical usage. To make the difference scores of natural and man-made stimuli comparable, the meaningful man-made characters were assigned to the shape level condition, while their designed inverted-mirrored counterparts were assigned to the form level.

In the Chinese writing system, some characters are mirror images of each other but have different meanings (e.g., "甲" and "由", "陪" and "部"). To eliminate the impact of these mirror words on character perception and to preserve the overall word form as much as possible, we first inverted the original words in both writing systems. We then mirrored the inverted text to create the form word stimuli. This transformation process is illustrated in Figure 2. Using novel inverted-mirrored words

as control stimuli has a significant advantage: both inverted and mirrored words are overlearned stimuli. Thus, participants' prior experience with typical instances of inverted or mirrored words could influence their perception of the experimentally created inverted-mirrored stimuli, especially if participants had no training on the modified versions. Moreover, the word length effect can be a confounding variable. The processing of a visual word is correlated with its length, usually expressed as the number of letters (Barton et al., 2014; New, Ferrand, Pallier, & Brysbaert, 2006). To control for this, all Chinese words used in this study were two characters long, and all English words were four letters long.

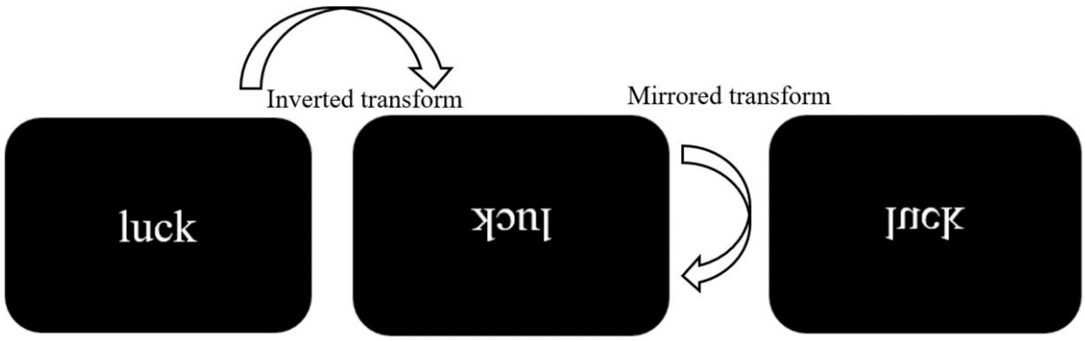


Figure 2. Schematic of the form words transformation process. In the second black screen, the word is inverted. Then the inverted text is mirrored, preserving the global word shape.

For the form condition of the man-made face stimuli—"Greeble"—we removed the face features using Photoshop software while keeping all other information the same as in the shape counterparts.

All 12 kinds of stimuli (six object categories with two levels) were shuffled. Participants were instructed to watch the screen and remain still. A total of 288 trials

were conducted, divided into two blocks with a 30-second rest. For each trial, the stimulus was presented for 500ms, followed by a fixation cross that appeared for a random duration between 900ms and 1100ms.

EEG recording

EEG was recorded continuously using a Neuroscan Grael amplifier (512Hz sampling rate: Cz reference) from 29 Ag/AgCl scalp electrodes mounted in an elastic cap and positioned according to the extended 10-20 system. The montage included 5 midline electrode sites (Fz, Cz, CPz, Pz, Oz) and 12 sites over each hemisphere (Fp1/Fp2, F3/F4, F7/F8, FC3/FC4, FT7/FT8, T7/T8, C3/C4, TP7/TP8, CP3/CP4, P3/P4, P7/P8, and O1/O2). Additional electrodes were used as ground, reference sites, and electrooculogram (EOG). Electrodes on the mastoids (M1/M2) were recorded but not used as a re-reference in this study.

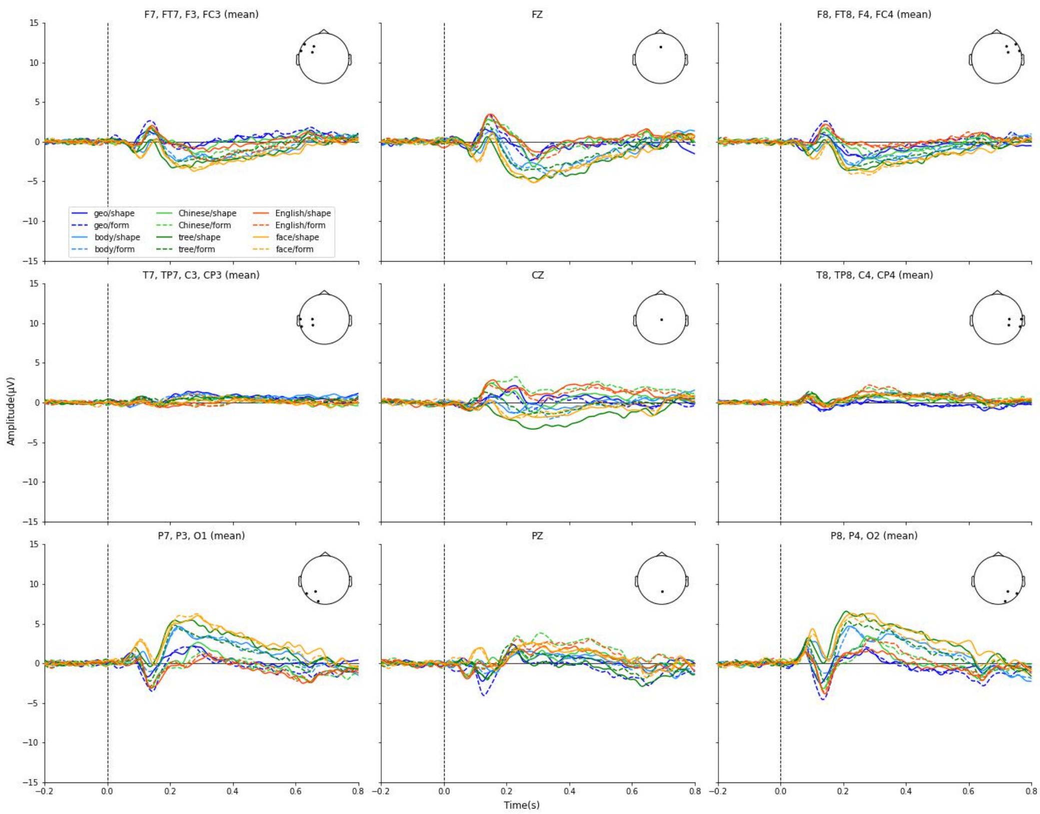
EEG preprocessing

All EEG data preprocessing was performed using MNE-Python software (Gramfort et al., 2013) v.1.5.1 in Python v. 3.12.0. The raw, continuous EEG data were bandpass-filtered in two ways: once using a highpass cut-off of 1Hz for artifact identification with independent components analysis (ICA) and once with a high-pass cut-off of 0.2Hz for further analysis. In both cases, the low-pass cut-off was 30Hz. ICA was applied to the continuous 1-30Hz bandpass-filtered data using the *fastica* algorithm (Hyvärinen, 1999), with the number of components set to explain 99% of the variance in the data. The ICA decomposition was then applied to the 0.2-30Hz bandpass-filtered data, and the EOG components were removed before further

analysis. The ICA-corrected data were segmented from 200ms prior to the onset of the stimuli to 800ms after. All segmented data were baseline-corrected and re-referenced to the average of all electrodes.

Results

The ERP grand waveforms elicited by the 12 object categories appeared generally similar, exhibiting a classic P1-N1-P2 pattern at lateral posterior electrode

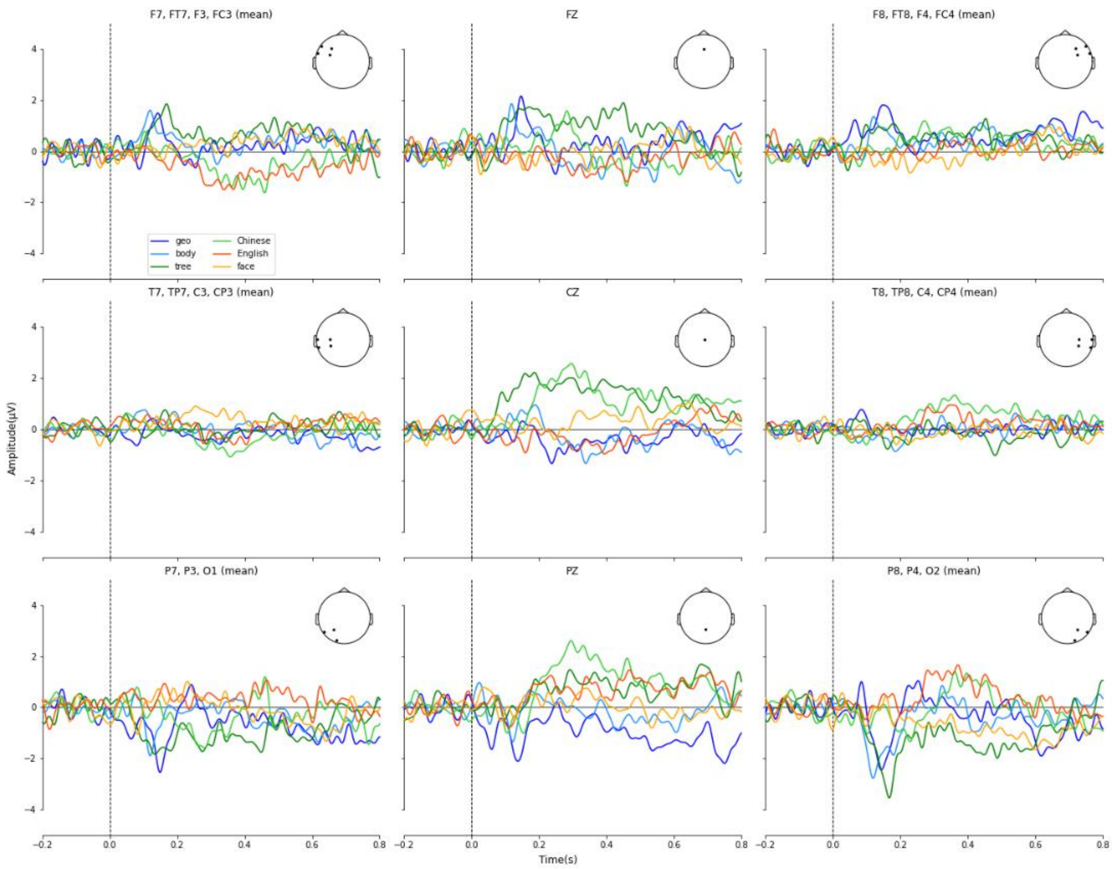


sites (**Figure 3**).

Figure 3. Grand average waveforms. The ERPs to shape (full lines) and form (dotted lines) for each condition at the nine ROIs. ERPs to stimuli in shape and form conditions are color-coded as follows: geometric figures in blue, animal bodies in dodgerblue, plants in green, Chinese characters in limegreen, English words in

orangered, and face in orange. Positive plotted upwards.

Since we were interested in the ERP differences between the shape and form condition level and their interaction with six types of object categories (the Target), we examined the difference waves created by subtracting the shape condition from the form condition for each target category. The difference waves are shown in **Figure 4**,



and the corresponding scalp topographic maps are presented in **Figure 5**.

Figure 4. Difference waveforms. The ERPs for the shape condition subtracted from those for the form condition are shown for different categories: geometric figures in blue, animal bodies in dodger blue, plants in green, Chinese characters in lime green, English words in orange-red, and Greebles in orange at the nine ROIs. Positive values are plotted upwards.

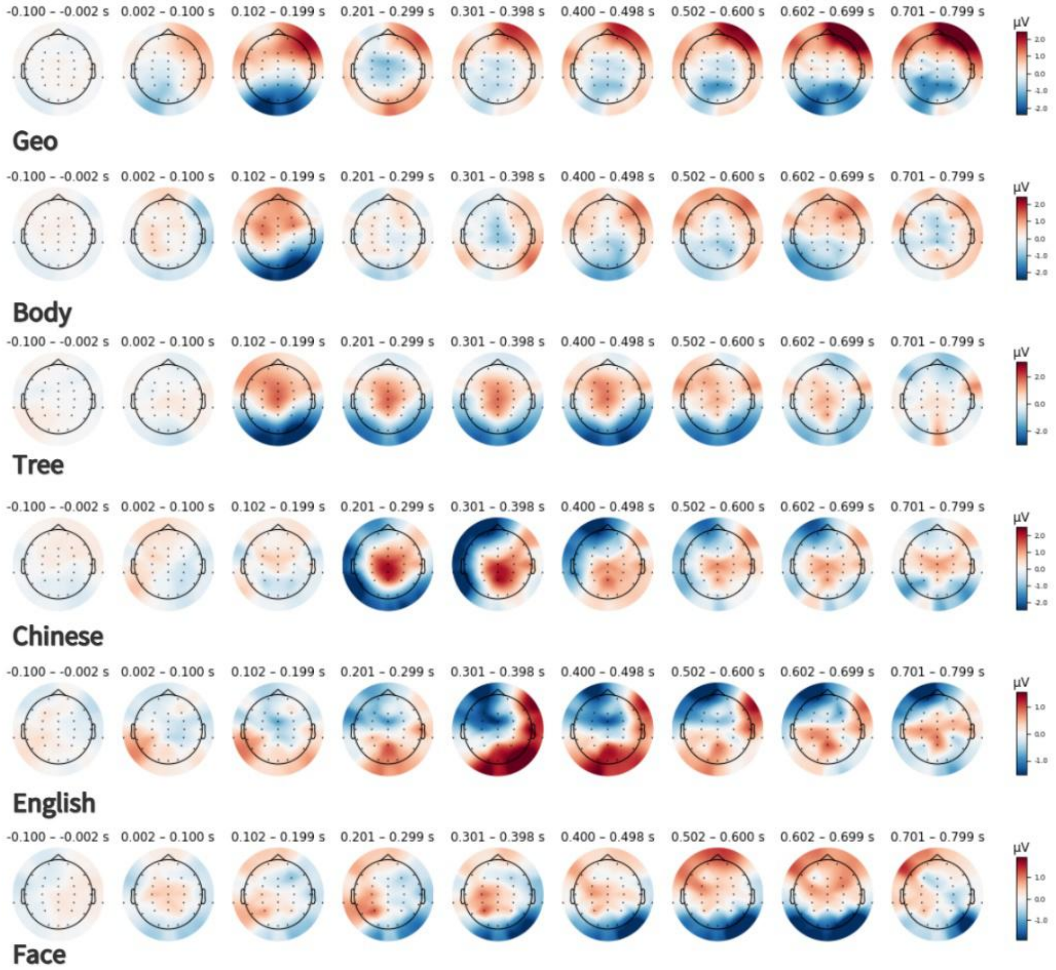
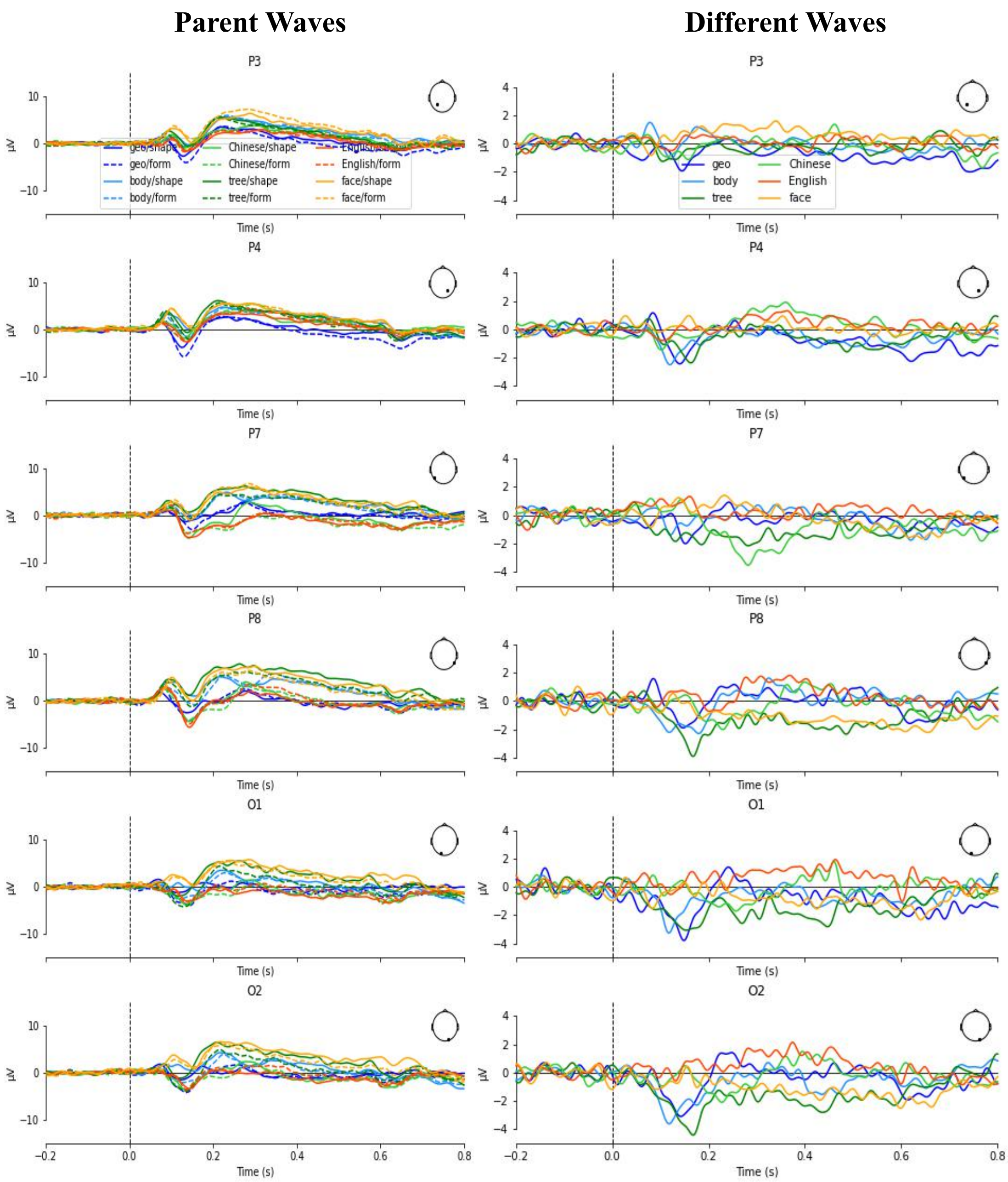


Figure 5. Topographic plots derived from different waveforms. The ERPs for the shape condition subtracted from those for the form condition are shown for each object category in the analyzed time windows (-100 to 0 ms,0 to100 ms,100 to 200 ms, 200 to 300 ms, 300 to 400 ms, 400 to 500 ms, 500 to 600 ms, 600 to 700 ms, 700 to 800 ms). Negativities are depicted in blue and positivities in amber.

The topographic map (**Figure 5**) showed that the first three conditions (geometric figures, animal bodies, and plants) shared similar difference waves during the period of 100-200ms after the onset of stimuli. Specifically, more negativity was observed over the vertex and extending to posterior sites from approximately 100-200ms, while more positivity appeared in the anterior sites. However, the last

three conditions (Chinese characters, English words, and Greebles) did not exhibit a common pattern. **Figure 4** suggested that the shared pattern was more obvious at the lateral posterior sites, especially for the right hemisphere. Indeed, a negative wave was shared among geometric figures (blue), animal bodies (dodger blue), and plants (green).

To pinpoint the specific channel of the shared ERP, we separately compared the parent waves with the different waves at the lateral posterior sites: P3, P4, P7, P8, O1,



and O2 (shown in **Figure 6**). The results indicated that the specific ERP was most profound at the O1 and O2 electrodes.

Figure 6. Parent waveforms and difference waveforms. The ERPs for geometric figures (blue), animal bodies (dodger blue), plants (green), Chinese characters (lime green), English words (orange red), and Greebles (orange) at the six channels: P3/P4, P7/P8, and O1/O2. Positive values are plotted upwards.

We computed mean amplitudes over a set of a priori time windows covering the post-stimuli period from 100-800 ms: 100-300 ms, 300-500 ms, 500-700 ms, 700-800 ms. The data were imported into the R software package v 4.3.0 and analyzed using the linear mixed effects modeling with the *bam* function in the *mgcv* package v 1.8-42 (Wood, 2017; Tremblay & Newman, 2015). Three candidate models were fit for each time window and compared using the Akaike information criterion (AIC; Akaike, 1973), which considers both the fitness and complexity of models. The best model for each time window included fixed effects of the target (geometric figures/animal bodies/plants/Chinese characters/English words/Greebles), condition levels (shape/form), and baseline, along with all possible interactions between these variables, as well as random intercepts for subjects (shown in **Table 1**). In all analyses, we excluded the interaction effect involving the baseline variable, meaning the baseline was controlled for in the results.

Table 1. Results of the LME model for Level, Target, and baseline at the O1/O2 electrodes.

	df	***F***	***p*-value**
100-300ms			

Level	1	4.48	0.0340
Target	5	63.86	0.0000
baseline	1	22.97	0.0000
Level:Target	5	8.92	0.0000
Level:baseline	1	4.33	0.0380
Target:baseline	5	0.39	0.8550
Level:Target:baseline	5	2.80	0.0160
300-500ms			
Level	1	0.39	0.5347
Target	5	24.18	0.0000
baseline	1	7.78	0.0053
Level:Target	5	9.72	0.0000
Level:baseline	1	0.43	0.5130
Target:baseline	5	0.77	0.5682
Level:Target:baseline	5	1.36	0.2355
500-700ms			
Level	1	6.23	0.0126
Target	5	11.35	0.0000
baseline	1	10.65	0.0011
Level:Target	5	3.82	0.0019
Level:baseline	1	0.20	0.6511
Target:baseline	5	1.21	0.2998
Level:Target:baseline	5	1.58	0.1624
700-800ms			
Level	1	5.10	0.0240
Target	5	2.02	0.0730
baseline	1	16.07	0.0001
Level:Target	5	1.48	0.1940
Level:baseline	1	0.29	0.5890
Target:baseline	5	1.98	0.0790
Level:Target:baseline	5	2.28	0.0440

Note: Level = shape and form; Target = geometric figures, animal bodies, plants, Chinese characters, English words, and Greebles.

Significant interactions involving Target and condition levels were further analyzed by comparing the form-shape contrasts for each target at the averaged ROI: O1 and O2 (shown in **Table 2.**). The *p*-values of these contrasts were corrected using the Benjamini and Hochberg (1995) method.

Table 2 Form-Shape contrasts.

Target	*t*	*p* (raw)	*p* (FDR BH)	Effect.Size	SE.eff	lower.CL	upper.CL
100-300ms							
geo	-2.1480	0.0318	0.0384	-0.1373	0.0639	-0.2627	-0.0120
bodies	-3.5154	0.0004	0.0013	-0.2236	0.0636	-0.3483	-0.0989
plants	-7.0476	0.0000	0.0000	-0.4499	0.0638	-0.5750	-0.3247
Chinese	-2.7285	0.0064	0.0128	-0.1736	0.0636	-0.2984	-0.0489
English	2.1447	0.0320	0.0384	0.1355	0.0632	0.0116	0.2593
Greebles	-1.5901	0.1119	0.1119	-0.1022	0.0643	-0.2282	0.0238
300-500ms							
geo	-0.6315	0.5277	0.6333	-0.0405	0.0641	-0.1663	0.0852
bodies	-0.8672	0.3859	0.5788	-0.0553	0.0638	-0.1804	0.0697
plants	-5.8397	0.0000	0.0000	-0.3755	0.0643	-0.5015	-0.2494
Chinese	0.3930	0.6944	0.6944	0.0251	0.0640	-0.1003	0.1506
English	3.5685	0.0004	0.0011	0.2274	0.0637	0.1025	0.3523
Greebles	-2.3778	0.0174	0.0349	-0.1542	0.0649	-0.2814	-0.0271
500-700ms							
geo	-2.5046	0.0123	0.0246	-0.1617	0.0646	-0.2883	-0.0351
bodies	-0.7939	0.4273	0.6410	-0.0509	0.0642	-0.1767	0.0748
plants	-4.1911	0.0000	0.0002	-0.2689	0.0642	-0.3948	-0.1431
Chinese	0.5011	0.6163	0.6862	0.0324	0.0646	-0.0943	0.1591
English	-0.4040	0.6862	0.6862	-0.0260	0.0644	-0.1522	0.1002
Greebles	-3.9787	0.0001	0.0002	-0.2557	0.0643	-0.3817	-0.1297
700-800ms							
geo	-2.2448	0.0248	0.1191	-0.1462	0.0651	-0.2739	-0.0185
bodies	0.8901	0.3734	0.7193	0.0576	0.0647	-0.0693	0.1845
plants	-0.3735	0.7088	0.7193	-0.0242	0.0648	-0.1513	0.1029
Chinese	-0.3595	0.7193	0.7193	-0.0234	0.0651	-0.1510	0.1042
English	-2.0575	0.0397	0.1191	-0.1331	0.0647	-0.2598	-0.0063
Greebles	-0.3802	0.7038	0.7193	-0.0247	0.0649	-0.1520	0.1026

100-300ms

Negativity was prominent in the difference waveforms and topographic maps, especially in the earliest time window: 100-200ms. Both the main effect of Level ($F(1) = 4.48$, $p = 0.0340$) and Target ($F(5) = 63.86$, $p < 0.0001$) were significant, and the interaction effect between Level and Target was also significant ($F(5) = 8.92$, $p < 0.0001$). The Form-Shape contrasts for each Target condition are shown in **Table 2**. The mean amplitude for geometric figures, animal bodies, plants, and Chinese characters showed significantly greater negativity in the form condition compared to

the shape condition at the averaged O1 and O2 electrodes, while the English words condition showed significant positivity in the form condition than the shape condition.

300-500ms

In this time window, the difference waveforms between form and shape were not significant ($F(1) = 0.39$, $p = 0.5347$). However, the main effect of Target ($F(5) = 24.18$, $p < 0.0001$) and the interaction effect between Level and Target were still significant ($F(5) = 9.72$, $p < 0.0001$). The Level×Target interaction was attributable to larger negativity for plants but larger positivity for English words.

500-700ms

In this time window, both the main effect of Level ($F(1) = 6.23$, $p = 0.0126$) and Target ($F(5) = 11.35$, $p < 0.0001$) were significant, and the interaction effect between Level and Target was also evident ($F(5) = 3.82$, $p = 0.0019$). Further contrast analysis showed that the geometric figures, plants, and faces categories exhibited greater negativity at the O1 and O2 channels in the form condition compared to the shape condition.

700-800ms

In the final time window, although the main effect of Level was still significant ($F(1) = 5.1$, $p = 0.0240$), which may have contributed to greater negativity in geometric figures and English words conditions, no significant difference between form and shape were found for each Target conditions under the corrected p values.

Discussion

To answer whether there is a specific brain signal corresponding to the form

analysis system, the present study designed form and shape conditions for various stimuli, including objects (geometric figures, still-standing animals without heads, and plants), words (mirror-inverted Chinese words and English words), and artificial face stimuli (the Greebles). We used the EEG method to probe adults' brain responses to these different kinds of stimuli. Our results are consistent with previous findings that different brain areas are involved in visual recognition when the stimuli vary among objects, words, and faces (Pegna et al., 2004; for reviews, see Farah, 1994).

We found a limited existence of form systems across object categories, including 2D geometric figures, animal bodies, and plants, but not in visual writing words or Greebles conditions. When comparing the parent waveforms in the form condition to those in the shape condition, a similar pattern of different waves was observed for 2D geometric figures, animal bodies, and plants across the O1 and O2 channels. These difference waves appeared between 100-200ms after stimuli onset and negatively peaked at the occipital lobe. This is the first time the form analysis system has been demonstrated across different types of object stimuli. Although the pattern seemed similar at first glance, adults exhibited a relatively long negative different wave for plant conditions. Compared with the other two object stimuli, the form condition in plant stimuli resulted in a longer period of negativity than the shape condition. The significant negativity waves for the former two conditions lasted from 100 ms to 300 ms, while the plant stimuli lasted from 100 ms to 700 ms. To our knowledge, no study has investigated human brain wave responses to plants. We were surprised by the long negativity response to plant stimuli but could not determine the underlying reason. Do

plants induce a more peaceful sense in humans, or do they hold a more ancient meaning compared with other kinds of object stimuli? Those questions remain unexplored.

The different waves of form-shape in the words condition lagged by 100 ms compared to the objects condition and exhibited a reverse pattern between the first and second language conditions. For Chinese participants, the Chinese words induced a negativity wave, while the English words elicited a positivity wave. During interviews, one participant reported that although the stimuli were transformed, the first-language words could still be recognized. In contrast, the second-language words could not be understood before disappearing. Thus, the significant positivity from 300 ms to 500 ms induced by the second language may reflect participants' surprise at not being able to figure out the meaning of the words (Lau, Phillips, Poeppel, 2008). It remains to be seen whether a similar effect exists in English-speaking participants.

The onset of the form-shape negativity wave of artificial face stimuli (Greebles) was not observed until 300ms. Most ERP studies described an earlier response to faces at around 170 ms, known as the N170, characterized by a vertex-positive and bilateral temporal-negative deflection (Bentin et al., 1996; Eimer, 2000; Pegna et al., 2002). The discrepant results may be due to the design of the form stimuli of Greeble. For the form condition of the man-made face stimuli, we removed the face features using Photoshop software while keeping all other information the same as in the shape counterparts. However, this operation removed the core form structure of face stimuli, such as eyes and mouth. Further studies should retain the facial features and

remove the geometric bodies of the Greeble stimuli.

The limited existence of the form analysis system in objects, but not in words and faces, may be due to the asymmetric characteristics of these stimuli. Indeed, the skeletons of the mentioned three types of object stimuli can be extracted by the form analysis system. Whether 2D geometric figures, animal bodies, or plants, all these stimuli's skeletons fit perfectly with the medial axis and grassfire illustration proposed by Blum. However, Blum's medial axis and grass fire model do not fit as well with word stimuli or face stimuli. Although we assume that words and faces also have their form or skeleton, their skeleton does not conform to the medial axis. This may be the core reason that the form analysis system is limited to objects and not applicable to words and faces.

However, the fact that our medial axis explanation, which fully explains our LO-oriented negativity waves limited to the object domain but not to the writing system or face stimuli, should not be over-interpreted. More experiments should be conducted before generalizing these results to a wider domain. According to Bao et al. (2020), at least four kinds of categories in two dimensions (animacy and spikiness) should be tested. Could this negativity wave also be observed in other domains that fit well with the medial axis model, such as man-made categories or silhouettes?

It should be noted that due to the low EEG spatial resolution, there was an unavoidable imprecision in localization. Thus, further studies using the fMRI method should pinpoint the precise brain areas activated by the form analysis system. Indeed, the present study suggests that the form analysis system exists in the lateral occipital

cortex; however, the further processing pathways to the temporal lobe or parietal lobe are still unclear. Traditional researchers believed that the ventral occipitotemporal pathway processes the properties of object perception, such as shape, texture, and color, whereas the dorsal occipitoparietal pathway encodes the spatial and temporal information of objects, such as motion and partial relations. Converging evidence, however, showed that both the ventral pathway and the dorsal pathway contributed to object recognition (Ayzenberg, Simmons, & Behrmann., 2023; for review, see Freud, Behrmann, & Snow., 2020). Recent studies have revealed that the dorsal pathway also represents the global shape (Ayzenberg and Behrmann., 2022). Moreover, considering both infants (Bertenthal, Proffitt, & Kramer, 1987) and adults (Johansson., 1950) are sensitive to the biological motion presented in point-light displays, and patients with dorsal pathway lesions but intact ventral pathways are selectively impaired in biological motion perception (Vaina et al., 1990), we suggest that the dorsal pathway also plays a pivotal role in form skeleton perception. Thus, further studies should investigate the corresponding roles of the ventral and dorsal pathways in the form analysis system.

This present study aimed to discover the neuro basis of the form analysis system of human vision, rather than exploring the developmental process that builds this system. How might biological development implement this process? Despite the tremendous variation in visual images, the form analysis system can represent the same object in different retinal positions, poses, distances, and sizes. Evidence from animals (Wood., 2013) and human infants (slater et al., 1990) shows that powerful,

robust, and invariant skeletal-based object recognition machinery is an inherent feature of the newborn brain. Thus, further steps could explore the neural basis of the form analysis system in newborns using EEG, to determine whether the form analysis system is core knowledge that happens before the onset of visual experience.

Finally, machine vision, analogy to human vision, remains one of the most challenging problems in artificial intelligence. Several studies compared AI-based deep convolutional neural networks to human vision and converging evidence suggested that convolutional network models do not classify based on global object shape as humans do (Baker, Lu, Erlikhman, & Kellman., 2018; 2020; Lowet, Firestone, & Scholl., 2018; Xu & Vaziri-Pashkam., 2021). For instance, when presented with a picture that shares the same shape as a cat but is filled with elephant texture, most people categorize this stimulus as a cat but all AI models classify it as an elephant. Our work may shed new light on machine vision. If a form analysis system exists in human vision, a new form analysis model could be developed and applied in the machine domain. Consequently, machine vision could one day see and categorize objects as effectively as human vision.

Conclusions

The present study revealed the existence of a form analysis system in the domain of object, including geometric figures, animal headless bodies, and plants, but not in different writing systems (i.e. Chinese and English) or Greebles. Shared negativity waves in the occipital lobe during the period of 100 ms to 200 ms were observed across the three object domains. These results can be perfectly explained by the

medial axis model, from which we can infer that the form analysis system may be applicable to a wider realm, as long as these categories fit well with the medial axis model.

References:

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle.

Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed letter confusability. *Journal of Memory and Language, 35*, 775-800.

Ayzenberg, V., & Behrmann, M. (2022). Does the brain's ventral visual pathway compute object shape?. *Trends in cognitive sciences*, *26*(12), 1119–1132. https://doi.org/10.1016/j.tics.2022.09.019

Ayzenberg, V., & Lourenco, S. F. (2019). Skeletal descriptions of shape provide unique perceptual information for object recognition. *Scientific Reports, 9*(9359), 1– 12.

Ayzenberg, V., Kamps, F. S., Dilks, D. D., & Lourenco, S. F. (2022). Skeletal representations of shape in the human visual cortex. *Neuropsychologia*, *164*, 108092. https://doi.org/10.1016/j.neuropsychologia.2021.108092

Ayzenberg, V., Simmons, C., & Behrmann, M. (2023). Temporal asymmetries and interactions between dorsal and ventral visual pathways during object recognition. *Cerebral cortex communications*, *4*(1), tgad003. https://doi.org/10.1093/texcom/tgad003

Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2018). Deep convolutional networks do not classify based on global object shape. *PLoS computational biology*, *14*(12), e1006613. https://doi.org/10.1371/journal.pcbi.1006613

Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2020). Local features and global shape information in object classification by deep convolutional neural networks. *Vision Research, 172,* 46–61. https://doi.org/10.1016/j.visres.2020.04.003

Bao, P., She, L., McGill, M., & Tsao, D. Y. (2020). A map of object space in primate inferotemporal cortex. *Nature*, 583, 103– 108.

Barton, J. J., Hanif, H. M., Eklinder Björnström, L., & Hills, C. (2014). The word-length effect in reading: a review. *Cognitive neuropsychology*, *31*(5-6), 378–412. https://doi.org/10.1080/02643294.2014.895314

Bentin S, Allison T, Puce A, Perez E, McCarthy G (1996): Electrophysiological studies of face perception in humans. *J Cogn Neurosci 8*:551–565.

Bertenthal, B. I., Proffitt, D. R., & Kramer, S. J. (1987). Perception of biomechanical motions by infants: implementation of various processing constraints. *Journal of experimental psychology. Human perception and performance*, *13*(4), 577–585. https://doi.org/10.1037//0096-1523.13.4.577

Blum H. (1973). Biological shape and visual science. I. *Journal of theoretical biology*, *38*(2), 205–287. https://doi.org/10.1016/0022-5193(73)90175-6

Bukach, C. M., Gauthier, I., Tarr, M. J., Kadlec, H., Barth, S., Ryan, E., Turpin, J., & Bub, D. N. (2012). Does acquisition of Greeble expertise in prosopagnosia rule out a domain-general deficit?. *Neuropsychologia*, *50*(2), 289–304. https://doi.org/10.1016/j.neuropsychologia.2011.11.023

Chambers, S. M. (1979). Letter and order information in lexical access. Journal of Verbal Learning and Behavior, 18, 225-241.

Cohen L, Dehaene S, Naccache L, Lehericy S, Dehaene-Lambertz G, Henaff MA, Michel F (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*,123:291–307.

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Fast mapping rapidly integrates information into existing memory networks. *Journal of experimental psychology. General*, *143*(6), 2296–2303. https://doi.org/10.1037/xge0000020

Cutting J. E. (1982). Two ecological perspectives: Gibson vs. Shaw and Turvey. *American Journal of Psychology*, 95, 199–222. Crossref. ISI.

Dehaene, S., Izard, V., Pica, P., & Spelke, E. (2006). Core knowledge of geometry in an Amazonian indigene group. *Science (New York, N.Y.)*, *311*(5759), 381–384.

https://doi.org/10.1126/science.1121739

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of experimental psychology. General*, *115*(2), 107–117. https://doi.org/10.1037//0096-3445.115.2.107

Duñabeitia, J. A., Molinaro, N., Laka, I., Estévez, A., & Carreiras, M. (2009). N250 effects for letter transpositions depend on lexicality: 'casual' or 'causal'?. *Neuroreport*, *20*(4), 381–387. https://doi.org/10.1097/wnr.0b013e3283249b1c

Eimer M (2000): The face-specific N170 component reflects late stages in the structural encoding of faces. Neuroreport 11:2319–2324.

Erdogan, G., & Jacobs, R. A. (2017). Visual shape perception as Bayesian inference of 3D object-centered shape representations. *Psychological review*, *124*(6), 740–761. https://doi.org/10.1037/rev0000086

Farah MJ (1994): Visual agnosia: disorders of object recognition and what they tell us about normal vision. Cambridge: MIT Press.

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(47), 18014–18019. https://doi.org/10.1073/pnas.0608811103

Firestone, C., & Scholl, B. J. (2014). “Please tap the shape, anywhere you like”: Shape skeletons in human vision revealed by an exceedingly simple measure. *Psychological Science, 25*(2), 377– 386.

Freud, E., Behrmann, M., & Snow, J. C. (2020). What Does Dorsal Cortex Contribute to Perception?. *Open mind : discoveries in cognitive science*, *4*, 40–56. https://doi.org/10.1162/opmi_a_00033

Freud, E., Behrmann, M., & Snow, J. C. (2020). What Does Dorsal Cortex Contribute to Perception?. *Open mind : discoveries in cognitive science*, *4*, 40–56. https://doi.org/10.1162/opmi_a_00033

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision research*, *37*(12), 1673–1682. https://doi.org/10.1016/s0042-6989(96)00286-6

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. *Journal of experimental psychology. Human perception and performance*, *28*(2), 431–446. https://doi.org/10.1037//0096-1523.28.2.431

Gauthier, I., Behrmann, M., & Tarr, M. J. (2004). Are Greebles like faces? Using the neuropsychological exception to test the rule. *Neuropsychologia*, *42*(14), 1961–1970. https://doi.org/10.1016/j.neuropsychologia.2004.04.025

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature neuroscience*, *2*(6), 568–573. https://doi.org/10.1038/9224

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "greeble" experts: A framework for studying expert object recognition processes. *Vision Research, 38*(15-16), 2401–2428. https://doi.org/10.1016/S0042-6989(97)00442-2

Grainger, J., & Whitney, C. (2004). Does the huamn mnid raed wrods as a wlohe? *Trends in Cognitive Sciences, 8*(2), 58–59. https://doi.org/10.1016/j.tics.2003.11.006

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-python. *Front. Neurosci.* 7:267. doi: 10.3389/fnins.2013.00267

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision research*, *41*(10-11), 1409–1422. https://doi.org/10.1016/s0042-6989(01)00073-6

Guerrera, C., & Forster, K. (2008). Masked form priming with extreme transposition. *Language and Cognitive Processes*, *23*(1), 117–142. https://doi.org/10.1080/01690960701579722

Gurariy, G., Mruczek, R. E. B., Snow, J. C., & Caplovitz, G. P. (2022). Using High-Density Electroencephalography to Explore Spatiotemporal Representations of Object Categories in Visual Cortex. *Journal of cognitive neuroscience*, *34*(6), 967–987. https://doi.org/10.1162/jocn_a_01845

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* 10, 626–634. doi: 10.1109/72.761722

Izard, Pica, P., & Spelke, E.S. (2021). Visual foundations of Euclidean geometry. *Cognitive Psychology, 136*.

Johansson, G. (1950). *Configurations in Event Perception.* Uppsala, Sweden: Almqvist & Wiksell.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Kao, C., Chen, D., & Chen, C. (2010). The inversion effect in visual word form processing. *Cortex, 46*, 217-230.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior research methods*, *44*(1), 287–304. https://doi.org/10.3758/s13428-011-0118-4

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS computational biology*, *10*(11), e1003915. https://doi.org/10.1371/journal.pcbi.1003915

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature reviews. Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Lowet, A. S., Firestone, C., & Scholl, B. J. (2018). Seeing structure: Shape skeletons modulate perceived similarity. *Attention, perception & psychophysics*, *80*(5), 1278–1289. https://doi.org/10.3758/s13414-017-1457-8

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: new evidence from the English Lexicon Project. *Psychonomic bulletin & review*, *13*(1), 45–52.

https://doi.org/10.3758/bf03193811

Pegna AJ, Khateb A, Murray MM, Landis T, Michel CM (2002): Neural processing of illusory and real contours revealed by high-density ERP mapping. Neuroreport 13:965–968.

Perea, M., Duñabeitia, J. A., & Carreiras, M. (2008). R34D1NG W0RD5 W1TH NUMB3R5. *Journal of experimental psychology. Human perception and performance*, *34*(1), 237–241. https://doi.org/10.1037/0096-1523.34.1.237

Psotka, J. (1978). Perceptual processes that may create stick figures and balance. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(1), 101– 111.

Rennig, J., Himmelbach, M., Huberle, E., Karnath, H.-O., 2015. Involvement of the TPJ area in processing of novel global forms. J. Cognit. Neurosci. 27, 1587–1600. https:// doi.org/10.1162/jocn_a_00809.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., & Linotte, S. et al., (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. *Neuroreport*, *11*(1), 69–74.

Schoonbaert, S., & Grainger, J. (2004). Letter position coding in printed word perception: Effects of repeated and transposed letters. *Language and Cognitive Processes, 19*(3), 333–367. https://doi.org/10.1080/01690960344000198

Slater, A., Mattock, A., & Brown, E. (1990). Size constancy at birth: newborn infants' responses to retinal and real size. *Journal of experimental child psychology*, *49*(2), 314–322. https://doi.org/10.1016/0022-0965(90)90061-c

Spelke, E. S. (2022). *What babies know: Core knowledge and composition.* Oxford University Press. https://doi.org/10.1093/oso/9780190618247.001.0001

Surtees, A. D. R., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of level - 2 perspective - taking in children and adults. *British Journal of Developmental Psychology, 30*(1), 75–86. https://doi.org/10.1111/j.2044-835X.2011.02063.x

Tanaka, J. W., & Farah, M. J. (1993). Parts and Wholes in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A*, *46*(2), 225-245. https://doi.org/10.1080/14640749308401045

Tanaka, J. W., & Simonyi, D. (2016). The "parts and wholes" of face recognition: A review of the literature. *Quarterly journal of experimental psychology (2006)*, *69*(10), 1876–1889. https://doi.org/10.1080/17470218.2016.1146780

Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception, 9*(4), 483–484. https://doi.org/10.1068/p090483

Tremblay, A., & Newman, A. J. (2015). Modeling nonlinear relationships in ERP data using mixed-effects regression with R examples. *Psychophysiology*, *52*(1), 124–139. https://doi.org/10.1111/psyp.12299

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*, 471–491. http://dx.doi.org/10.1111/j.2044- 8295.1988.tb02747.x.

Vaina, L. M., Lemay, M., Bienfang, D. C., Choi, A. Y., & Nakayama, K. (1990). Intact "biological motion" and "structure from motion" perception in a patient with impaired motion mechanisms: A case study. *Visual Neuroscience, 5*(4), 353–369. https://doi.org/10.1017/S0952523800000444

Wandell, B. A., Rauschecker, A. M., & Yeatman, J. D. (2012). Learning to see words. *Annual review of psychology*, *63*, 31–53. https://doi.org/10.1146/annurev-psych-120710-100434

Wilder J, Rezanejad M, Dickinson S, Siddiqi K, Jepson A, & Walther D. (2019). Local contour symmetry facilitates scene categorization. *Cognition, 182*:307-317. DOI: 10.1016/j.cognition.2018.09.014.

Wood J. N. (2013). Newborn chickens generate invariant object representations at the onset of visual object experience. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(34), 14000–14005. https://doi.org/10.1073/pnas.1308246110

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R.* 2nd Edn. Chapman and Hall/CRC.

Xiao, Q., Qin, M., & Yin, Y. (2020). Skeleton-based Chinese sign language recognition and generation for bidirectional communication between deaf and hearing people. *Neural networks : the official journal of the International Neural Network Society*, *125*, 41–55. https://doi.org/10.1016/j.neunet.2020.01.030

Xu, Y., & Vaziri-Pashkam, M. (2021). Limits to visual representational correspondence between convolutional neural networks and the human brain. *Nature communications*, *12*(1), 2065. https://doi.org/10.1038/s41467-021-22244-7

Yin, R. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141–145.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*(6), 747–759. https://doi.org/10.1068/p160747

Zhang, W., Ni, A., & Li, S. (2021). Development of orientation sensitivity to Chinese compound characters in 4- to 6-year- old children. *PsyCh journal*, *10*(1), 33–46. https://doi.org/10.1002/pchj.405